Pergamon

0360-1315(95)00078-X

Computers Educ. Vol. 26, No. 1–3, pp. 33–40, 1996

0360-1315/96 $15.00 + 0.00

ARE INTEGRATED LEARNING SYSTEMS EFFECTIVE LEARNING SUPPORT TOOLS?

J. UNDERWOOD,[1,2] S. CAVENDISH,[1,2] S. DOWLING,[2] K. FOGELMAN[2] and T. LAWSON[2]

[1] ESRC Centre for Research in Development, Instruction and Training and [2] School of Education, Leicester University, 21 University Road, Leicester LE1 7RF, England

Abstract—The following summary presents the key findings of an evaluation of a 6 month trial of one Integrated Learning System (ILS) in nine U.K. Schools, conducted by the School of Education, Leicester University. The pupils in the target population ranged from 8 to 13 years of age. The evaluation design was based on the comparison of outcome performance in mathematics (basic numeracy) and reading for designated experimental and control groups in five schools, although additional data from all nine schools enriched this central core of data. In addition the evaluators monitored pupil behaviour. There were significant learning gains in the ILS treatment group for mathematics (basic numeracy) compared to the Control group. However, no such gains were recorded for reading. There were no discernible differences in patterns of attendance at school for ILS and Control groups. In general, behaviour was good when using the ILS with higher time on task and a calmer working atmosphere than in the standard classroom. There was some indication that this improved behaviour transferred to the standard classroom over time.

CONTEXT OF THIS INVESTIGATION

One outcome of the 1993 visit by members of the Department for Education (DfE) and the National Council for Educational Technology (NCET) to the U.S.A. was an increased interest in Integrated Learning Systems (ILSs). Such systems have been in use in the U.S.A. for a number of years and there is some evidence that they are an effective learning mode, although the findings do need to be viewed with care [1].

In the spring and summer terms of 1994, NCET conducted a national trial of two software packages in twelve U.K. schools. The purpose of the trial was to assess whether the use of ILSs would prove to be an appropriate and effective approach to learning in a U.K. context. Central to that focus was an evaluation of children's acquisition of basic skills in numeracy and literacy. This summative evaluation of learning gains was conducted by the team at Leicester University. The research presented here is the result of our 6 month evaluation of one of those ILSs, Computer Curriculum Corporation's (CCC) *SuccessMaker*—version 15.

What is an integrated learning system?

The term Integrated Learning System (ILS) is an ill-defined and often misused term. It derives from the U.S. and is held to mean a system that includes extensive courseware plus management software usually running on a networked system [2]. There are three essential components to such a system: the curriculum content, the record system, and the management system. Much of the domain-specific educational software in the U.K. incorporates the first two components: that is a range of tutorial materials and assessment modules linked to a student record system. One dividing line between an ILS and many of these tutoring packages is that of scale. An ILS will have substantial course content and aggregated learner records. A more fundamental difference between such packages, however, is the management system of the ILS. The functionality of the management system may vary but at a basic level it will update students' records, interpret learner responses to the task in hand and provide performance feedback to the learner and the teacher. These data collected and interpreted by the system may also be used to provide an 'individualized' pathway through the curriculum content. The degree to which the locus of control over individual learning sequences lies with the machine, the teacher or the learner is a key difference between the various systems now on the market.

CCC *SuccessMaker*, the software at the heart of this evaluation, is generally viewed as a closed

Table 1. The distribution of pupils in the ILS and Control groups

Schools	Pilot	Control
CCC Primary	126	27
CCC Secondary	125	104
Total	251	131

Table 2. The distribution of pupils by sex in the ILS and Control groups

Schools	Boys	Girls
CCC Primary	71	82
CCC Secondary	130	99
Total	201	181

system, i.e. the curriculum content and the learning sequences are not designed to be changed or added to by either the tutor or the learner. This does not mean that all learners will take the same pathway through the material. The management system, within the bounds of pre-set algorithms, will produce a differentiated teaching sequence for each learner based on his or her previous and current performances.

Purpose of the evaluation

The key objectives of the summative evaluation conducted by Leicester University were to assess those learning gains within basic numeracy and literacy that could be attributed to the use of the ILS; to identify the effects of individual differences on learning outcomes; and, where possible, to illuminate those features of ILS which supported learning.

METHOD

The sample

The sample schools. Nine sample schools took part in the project, one school each in Northern Ireland, Wales and Scotland, and the remaining 6 schools distributed throughout England. There were four primary schools and five secondary schools. Each school used the *SuccessMaker* software for mathematics and English. Although curriculum software areas other than mathematics and English were available to the schools, work in such areas did not form part of the external evaluation.

The pilot sample. The pilot sample was a sub-set of the pupils working on the ILS. The pilot group constituted approximately one-third of the total students working within the full NCET project. Eight out of nine pilot schools were situated in areas of predominantly council owned property. Local unemployment was high for all but two schools ranging from 30 to 80% with 50% free school meals being quite normal. Two schools had significant ethnic minority populations.

Table 1 shows the number of ILS and Control group pupils involved in the evaluation study. The age range was from 8 to 13 years. Although both younger and older pupils were involved in the NCET ILS project, the formal evaluation focused on pupils at Key Stages Two and Three.

There were comparable numbers of boys to girls involved in the evaluation study (Table 2). There was one primary girls' school and one secondary boys' school in the sample.

The evaluation design

A summative evaluation model was used to investigate the impact of the ILS treatment on children's learning. The outcome measures were numeracy and reading gains. Five schools agreed to allow the evaluators to test not only those pupils who were to use the ILS system but also a

Table 3. Overall mean mathematics and reading scores for the collective ILS and Control pupils

Set	Pre-trial	Post-trial
Maths all Primary pupils	39.4	46.6
Maths all Secondary pupils	29.1	37.4
Reading all Primary pupils	9.8	10.0
Reading all Secondary pupils	10.8	11.1

Control group of peers who would continue working in their standard classrooms. The need for control groups is convincingly argued by Becker in his critique of previous evaluations [1]. The evaluation team was unable to select control groups but rather were faced with a limited choice of schools which could provide us with at least two comparable age and ability groups within the school.

All children in both the pilot ILS groups and the Control groups completed standardized tests of non-verbal ability [3], mathematical achievement [4, 5] and reading achievement [6, 7] at the start (pre-test) and at the end (post-test) of the 6 month trial period. In addition termly aggregated school attendance figures were taken over the period of the evaluation for each of the ILS and Control classes. Pupil attitude to core areas of study formed a further pre–post-measure.

The pre–post-test data were supplemented by observations of both the whole class and targeted individual pupils in their standard classroom activities immediately pre- and post- their ILS session and working on the ILS; and subject and class teachers' comments on selected students' attendance, effort and achievement over the period of the evaluation.

RESULTS

Performance outcomes

The baseline. A correlation analysis confirmed predicted positive relationships between non-verbal reasoning, mathematical and reading performances at both pre- and post-test level for both Primary and Secondary pupils.

Termly attendance figures for all pilot pupils were recorded throughout the year. The pre-study Autumn term attendance provided the evaluators with a base line measure prior to the intervention treatment (i.e. prior to the start of ILS use). Termly attendance levels taken pre, mid and end of the trial are generally positively correlated but only the level of attendance at the Secondary schools in the spring and summer terms showed any relationship to performance. In these schools post-trial mathematics performance is positively related to attendance. Autumn term attendances acted as a base line measure for the ILS and Control group behaviours in the spring and summer terms.

Learning gains. For both Primary and Secondary there was an improvement in mathematics and reading performance over the trial period for all pupils (Table 3).

Treatment effect. Were the learning gains differentially attributable to the use of ILS? The natural definition of change is the difference between initial and final scores but such a simple measure is unsatisfactory for a number of reasons [8]. Here we have chosen to perform an analysis of co-variance to ascertain whether or not the post-trial mathematics and the post-trial reading scores for the ILS pupils are significantly different from those of the Control pupils. Although every effort was made to match ILS and Control pupils at the start of the evaluation process, there were inevitable differences between the groups. It was an assumption of this study that the pre-trial performance in both mathematics and in reading, and non-verbal reasoning and chronological age, would influence the relevant post-test performance. In order to account for their influence these factors were entered as independent variables into the analysis.

Secondly, in order to remove, or at least reduce, the effect of variables such as 'social/home background' in the analysis only schools providing both ILS and Control pupils were entered into the comparative group analysis. The sample size is therefore reduced.

Table 4. Mean adjusted post-trial performance scores for mathematics for Primary and Secondary CCC schools

Set	ILS mean % maths score	Control mean % maths score	Significance level of difference
Primary CCC	48.31	39.16	$P<0.001$
Secondary CCC	40.21	34.77	$P<0.001$

Table 5. Summary of effect size analysis for Primary and Secondary CCC schools

Set	Effect size mathematics
Primary CCC	+0.4
Secondary CCC	+0.4

Effect Size: the difference in performance between the ILS and Control group, expressed as a proportion of 1 SD of pre-trial scores for the combined ILS and Control groups (after Becker [1]).

Performance gains in mathematics. The mean score differences, adjusted as described above, for the ILS and Control groups are presented in Table 4. To help the reader's understanding of the data, scores are presented as a percentage of the total possible score on the post-test. For Primary, the ILS group ($N=20$) performed significantly better than the Control group ($N=19$), achieving an average post-trial mathematics score that was 9.15% higher than the Control group.

Further analysis by mathematics sub-skills showed that the ILS group had superior performances in Addition ($F=20.7$, d.f. = 1, 34, $P<0.001$); Subtraction ($F=7.89.7$, d.f. = 1, 34, $P<0.008$); Multiplication ($F=6.29$, d.f. = 1, 34, $P<0.02$); and Extensions ($F=7.9$, d.f. = 1, 34, $P<0.008$) but that there were no significant differences between the groups for Division nor for Measurement and Money.

For Secondary, the ILS group ($N=77$) performed significantly better than the Control group ($N=57$), achieving an average post-trial mathematics score that was 5.44% higher than the Control group (Table 4).

Further analysis by mathematics sub-skills showed that the ILS group had superior performances in Operations ($F=4.85$, d.f. = 1, 118, $P<0.03$) and Diagrams ($F=5.41$, d.f. = 1, 118, $P<0.02$); but that there were no significant differences between the groups for Addition, Subtraction, Multiplication, Division, Extensions, nor for Measurement and Money. Performance on Fractions was better for the ILS group but failed to reach the $P<0.05$ significance level.

In the secondary school analyses it was possible to include 'school' as an additional independent variable in order to check for the presence of interactions between school and teaching strategy, that is whether the effects of the latter were similar in different schools. For the Secondary schools there were no significant interactions. The pattern of mathematics gains was consistent over all schools.

Significance of the gains. How important are the differences that have been identified? Becker has argued that it is essential to present the findings of an evaluation such as this in an internationally recognized form. He argues that the common measure of effectiveness of a treatment such as teaching styles is the Effect Size (E.S.). The E.S. for this current study is the difference in performance between the ILS and Control group, expressed as a proportion of one standard deviation (SD) of pre-trial scores for the combined ILS and Control groups [1]. Becker argues that an effect size of 0.15 over a year is negligible as it can be accounted for, among other things, by maturation. An E.S. above 0.15 is significant however as it can not be accounted for by children's variation in natural increase in performance with age.

The E.S. (Table 5) for Primary and Secondary mathematics was +0.4 indicating a substantial positive effect on pupil achievement.

Table 6. Mean adjusted post-trial performance scores for reading for Primary and Secondary CCC schools

Set	ILS mean reading age	Control mean reading age	Significance level of difference
Primary CCC schools	10.23	10.29	NS
Secondary CCC schools	11.13	11.26	NS

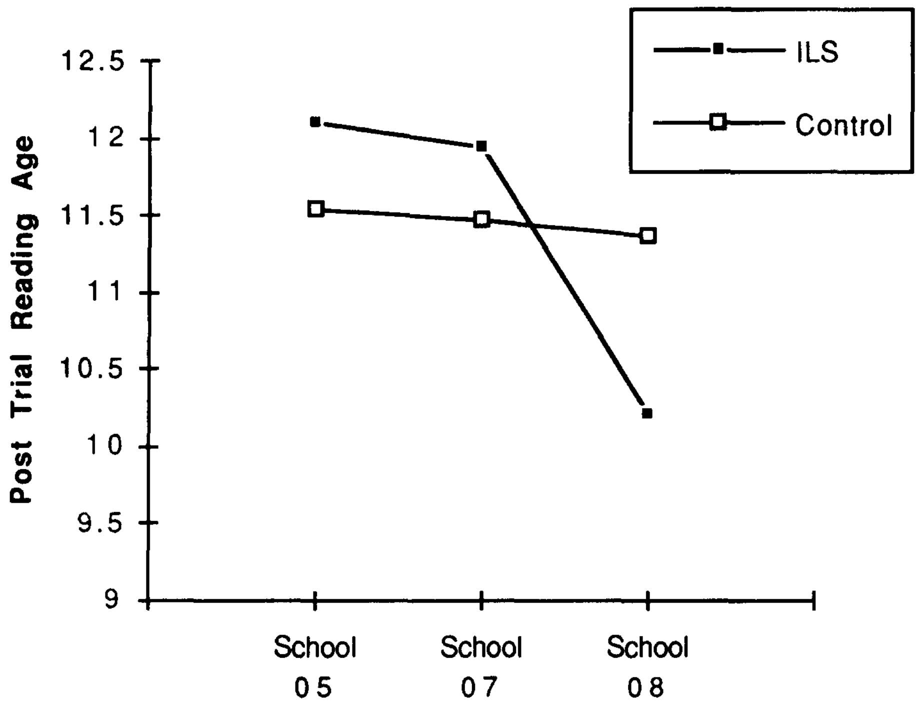

Fig. 1. Secondary CCC post-trial adjusted reading ages by teaching strategy and school.

Performance gains in reading. As for mathematics, a comparison of ILS and Control group reading performance was conducted using analysis of covariance. The scores in Table 6 are presented as reading ages, as provided by the test standardization data.

There was no significant difference between the ILS and Control groups (Table 6) at either primary or secondary level but there was a two way interaction between teaching strategy and school ($F = 2.97$, d.f. = 3, 146, $P < 0.05$). The pattern of performance presented in Fig. 1 is confused. The ILS group in two schools perform better than the Control groups but in the third school the position is reversed and the Control group out-performed the ILS group. The interaction reported here was the only point at which the ILS and Control groups could be differentiated in their reading performance.

Additional evidence of reading performance was supplied by two schools. Schools 03 and 09 conducted their own pre- and post-trial reading tests. School 03 had no Control pupils in their sample but School 09 tested both Control and ILS pupils.

Although School 09 has recorded large post-trial performance gains in reading these data are uninterpretable as the children did not complete the same test pre- and post- the project.

The test used by School 03 had a multiple choice format which was similar to that of the ILS. It might be predicted therefore that such a test would record higher performance post-test gains than the CLOZE type test used by the evaluators. This proved to be so but although the gains in reading age were higher than those from the formal evaluation, the ranking of children remained the same. Pre–post-gains by these pupils, although larger, were of the order of half a year's improvement over the 6 month trial period. This supports the formal evaluation's finding that there were no major gains in reading. The American studies, however, would predict that gains

Table 7. Mean adjusted pre-trial test scores in mathematics and reading by levels of gain for ILS and Control pupils

	Low gain pupils	Average gain pupils	High gain pupils	Significance level of difference
CCC schools maths	16.01	29.35	42.83	$P<0.001$
CCC schools reading	11.39	10.47	10.25	$P<0.001$

in reading would be small and probably not discernible after only 6 months. We feel it is too early to make any definitive comments about the impact of ILSs on reading.

Individual differences

A key question for this study was to establish whether identifiable groups of pupils were differentially affected by their time on the ILS. There were no significant age or gender differences in performance gains.

Three groups of pupils were identified, pupils who made low gains, pupils with average gains and pupils with high gains, for both mathematics and reading. To allow comparisons across age groups the overall performance gain (the difference between the pre- and post-test scores) was computed for the ILS and the Control pupils together. The groups were formed by dividing this statistic into upper, middle and lower thirds. The data were inspected by analysis of co-variance. Separate analyses were conducted for schools in mathematics and in reading. Mean adjusted scores of performance on the pre-tests are presented in Table 7.

In reading, the low gains group ($N=68$) had significantly higher pre-test scores than the average ($N=51$) and above average groups ($N=74$). The pupils who made most gains in reading then, were those who had a lower starting point. This was true for both ILS and Control pupils, and so the ILS software did not benefit one achievement group over another to a greater extent than one would find with standard classroom teaching.

In mathematics, the low gains group ($N=50$) had significantly lower pre-test scores than the average ($N=61$) and above average gains groups ($N=58$). Their adjusted average starting score was 27% lower than their most able peers (Table 7). The pupils who made most gains in mathematics, were those who had a higher starting point. This was true for both ILS and Control pupils and so the ILS software did not benefit one achievement group over another any more than one finds with standard classroom teaching.

The failure of the lowest achievers to make substantial learning gains from the system concurs with the general observations of both teachers and the evaluators. Although many teachers had an initial expectation that the least able would benefit most from working with the ILS, observation of their pupils using the system showed that this was unlikely to be so. There are a number of possible reasons for this. Certainly these children found it very difficult to maintain their span of attention and a number of schools have argued for a shorter contact time with the system. Anecdotal data also suggested that the lower ability children were prone to guessing when faced with multiple choice type questions.

In our interviews with children a number expressed a dislike of multiple choice questions because they were 'cheating' or 'not proper work'. Other pupils however liked this type of question because it reduced the amount of writing!

Behaviour patterns on and off the ILS

General impressions of whole class behaviour were obtained by observation of the whole group during lessons immediately prior to the group's ILS session, during the ILS session, and for the lesson immediately following the ILS session. In addition to whole class observation, individual pupils were targeted for in-depth study.

We recorded high time on task and a calm atmosphere which was appreciated by children. Pupils with behaviour problems were able to concentrate more than in standard classes although they did not maintain those good behaviour patterns for all sessions.

In general children worked constructively on the ILS for about 15 minutes at a time before requiring a change of activity. Lower ability levels may find even this length of time beyond their attention span. Primary children had high attention and good behaviour patterns during the ILS sessions but showed some deterioration of behaviour over time for post-ILS lessons. Pupils with behaviour problems were able to concentrate more than in standard classes although they may not maintain those good behaviour patterns for all of the ILS sessions.

Secondary pupils showed high levels of attention and good behaviour patterns during ILS sessions. This improved behaviour carried over into the standard classroom. Subject teacher assessment of targeted ILS and Control pupils' behaviour and performance was recorded through a short questionnaire at the mid-term and at the end of the project. At the end of the project, the targeted ILS pupils' performances were rated more highly than the Control target children by their subject teachers ($F = 9.13$, d.f. = 1, 90, $P < 0.003$). The effect size (E.S.) was +0.39. Also taking the first reports as a baseline for data analysis, the teacher assessments indicated that the level of improvement in pupils' achievement throughout the trial period was greater for the ILS than for the Control pupils, with an effect size of +0.36 ($F = 4.16$, d.f. = 1, 90, $P < 0.044$).

There were no discernible differences in patterns of attendance for ILS and Control groups.

SUMMARY CONCLUSIONS

Children working on the CCC ILS performed significantly better in mathematics than children working in the Control groups. The effect size was substantial (+0.4). Taking the Control group's progress as a baseline for 6 months' growth, the CCC ILS mathematics group made the equivalent of up to 20 months growth in that 6 months' period. This finding concurs with CCC's own studies and with Becker's evaluation of a second commercial system [1].

There was no discernible difference between the ILS and Control reading groups. Becker's review of the evidence from American schools would suggest that a small effect might be found over a full year. The lack of any finding over a 6 month period is therefore not surprising. The question of the effectiveness of the CCC ILS for language development remains open.

In terms of behaviour there are some preliminary indications that good work practices are being transferred from the ILS classroom to the standard classroom.

DISCUSSION

Many, both within and outside of the education profession, are uncomfortable with the concept of ILS. Wragg clearly articulates their discomfort "the majority of parents do not want their children to learn by sitting at a screen all day long. Most normal adult behaviour is interactive—and you get some strange people if they've been up in the attic starting at a computer screen from the age of 5 to 16" (p.11) [9]. This should not be an issue for ILS. The evaluation team found that approximately 30 minutes a day was the optimal time on the system [10]. More lengthy stays at the machine lead to reduced motivation and poor behaviour. Pea [11] goes further in questioning the validity of the underpinning goal of such an education whose central goal is to use technology to 'enhance' or 'make more efficient' the delivery of key components of instructional activity inherent in a traditional education. Such an education is based on the concept of solo intelligence and not on that of distributed intelligence.

The findings presented here, however, suggest that the ILS approach to teaching and learning should be taken seriously. The gains within basic numeracy were significant and the reading performance did not suffer from the use of an ILS approach. We are not arguing that major policy decisions should be made on the basis of these data rather that further research into such systems is warranted. There are questions that are still to be answered. For example, are the gains we recorded sustainable over time or are they a quick fix. That is, do those children who are performing significantly better than their peers maintain their advantage or is there a ceiling effect. If such children stop using the system, do their performances and those of their peers re-converge? Our sample population tended not to be drawn from more academic schools. We still need to know whether children from such schools will benefit from the use of such systems. Equally we need to know whether this style of learning is acceptable for all children.

On a different note we need to know why ILSs are effective. Are all such systems likely to be effective or are there particular aspects of systems that stimulate learning? *Success Maker* appears to operate contingently, scaffolding the learning, i.e. it provides help as the child needs it through its tutorial OWL and focuses the learners' attention on displayed weaknesses in their knowledge. Is such a system always of benefit or can children become prisoners of this structured help?

In Phase 2 of our evaluation which is again funded by NCET under the auspices of the DfE we hope to begin to answer some of these extended questions.

Acknowledgements—To all the staff and pupils of the participating schools for their time, effort and co-operation. To our consultants, Professor Geoffrey Underwood and Professor David Wood, Department of Psychology, University of Nottingham and the ESRC Centre for Research in Development, Instruction and Training. To Professor Maurice Galton for his helpful comments. This is an NCET/DfE funded project.

REFERENCES

1. Becker H. J., Computer-based ILSs in the elementary and middle grades: a critical review and synthesis of evaluation reports. *J. Educ. Comput. Res.* **8,** 1–42 (1992).
2. Office of Technology Assessment, *Power on! New Tools for Teaching and Learning.* U.S. Government Printing Office, Washington, D.C. (1988).
3. Ravens J., *Standard Progressive Matrices.* Lewis, London (1956).
4. France N., *Profile of Mathematical Skills: Level 1.* NFER-Nelson, Windsor (1979).
5. France N., *Profile of Mathematical Skills: Level 2.* NFER-Nelson, Windsor (1979).
6. Macleod J., *GAP Reading Comprehension.* Heinemann, London (1970).
7. Macleod J., *GAPADOL Reading Comprehension.* Heinemann, London (1970).
8. Youngman M., *Analysing Social and Educational Research Data.* McGraw–Hill, London (1979).
9. Picardie J., Have traditional Schools had their day? *The Independent on Sunday Review* 2 April (1995).
10. Underwood J., Cavendish S., Dowling S., Fogelman K. and Lawson T., Integrated learning systems in UK schools. NCET, Coventry (1994).
11. Pea R., Practices of distributed intelligence and designs for education. In *Distributed Cognitions* (Edited by Salomon). Cambridge University Press, New York (1993).